EPIDPP36/165

Colorectal Cancer And *Streptococcus Gallolyticus* Infection: Accuracy Of Statistical And Machine Learning Models For Early Detection Algorithm

Edre Mohammad Aidid[1], Hairul Aini Hamzah[1], Mohd Shaiful Ehsan Shalihin[1], Azmi Md Nor[1], Che Muhammad Khairul Hisyam Ismail[1]
[1]Kulliyyah of Medicine, International Islamic University Malaysia, Kuantan Campus, Pahang, Malaysia

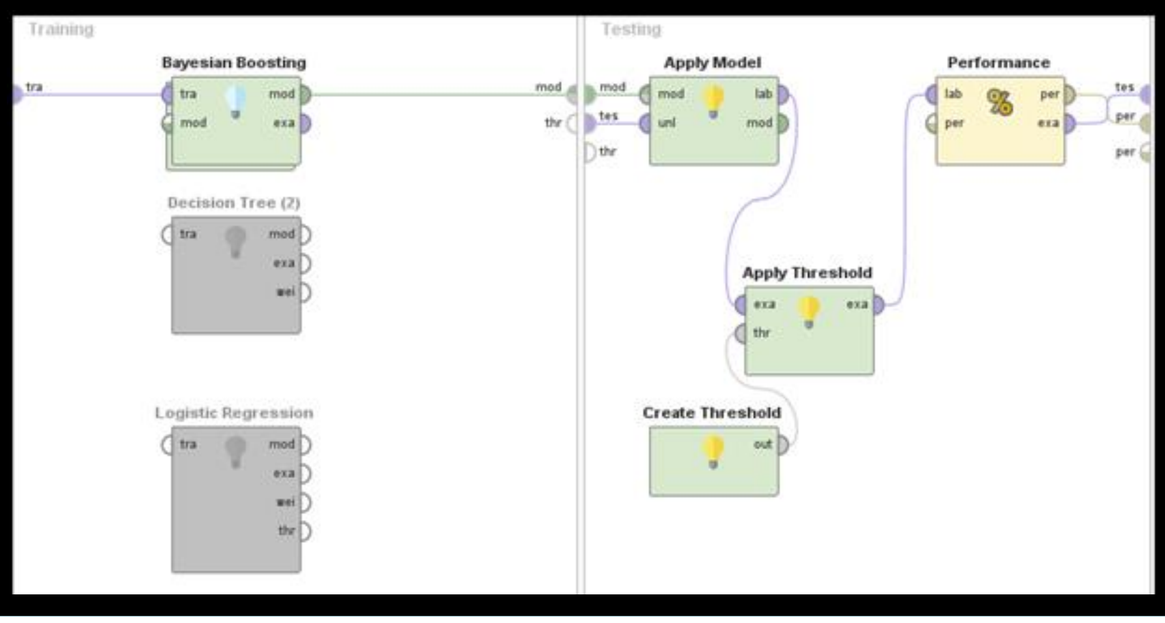
INTRODUCTION

1. **Epidemiological studies highlighted *Streptococcus gallolyticus* (SGG) infection in colorectal cancer (CRC) but was given less attention as compared to stool occult blood test (iFOBT)(1,2).**
2. **ML has been used to predict cancers but not as specific and as common as multivariate statistical models (3).**
3. **The research aimed to determine diagnostic accuracy of simultaneous testing of stool polymerase chain reaction (PCR) for SGG and iFOBT in detecting CRC and to compare accuracy of statistical and machine learning (ML) models.**

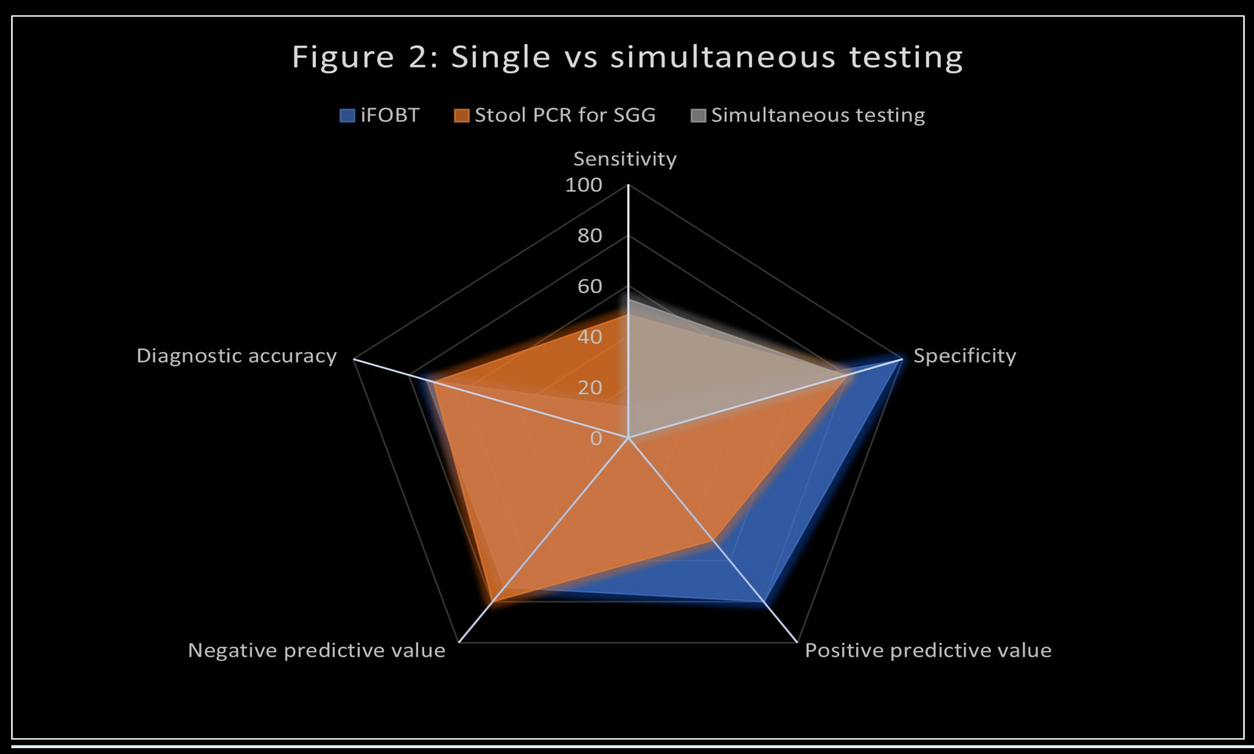
MATERIALS AND METHODS

1. **Case control study (1 case: 2 controls) 2019-2022**
   - **Confirmed CRC as the cases with controls coming from the same surgery clinic in SASMEC@IIUM.**
   - **Total sample size calculated was 120.**
   - **Modified Asia Pacific Colorectal Screening (APCS) score used for patient risk stratification.**
2. **Data analysis**
   - **Statistical – Logistic Regression (LR)**
   - **ML – Decision tree, Bayesian decision tree (BDT)**

Figure 1 : Cross validation was utilized for the comparative analysis using Rapidminer studio version 9.10.011



RESULTS AND DISCUSSION



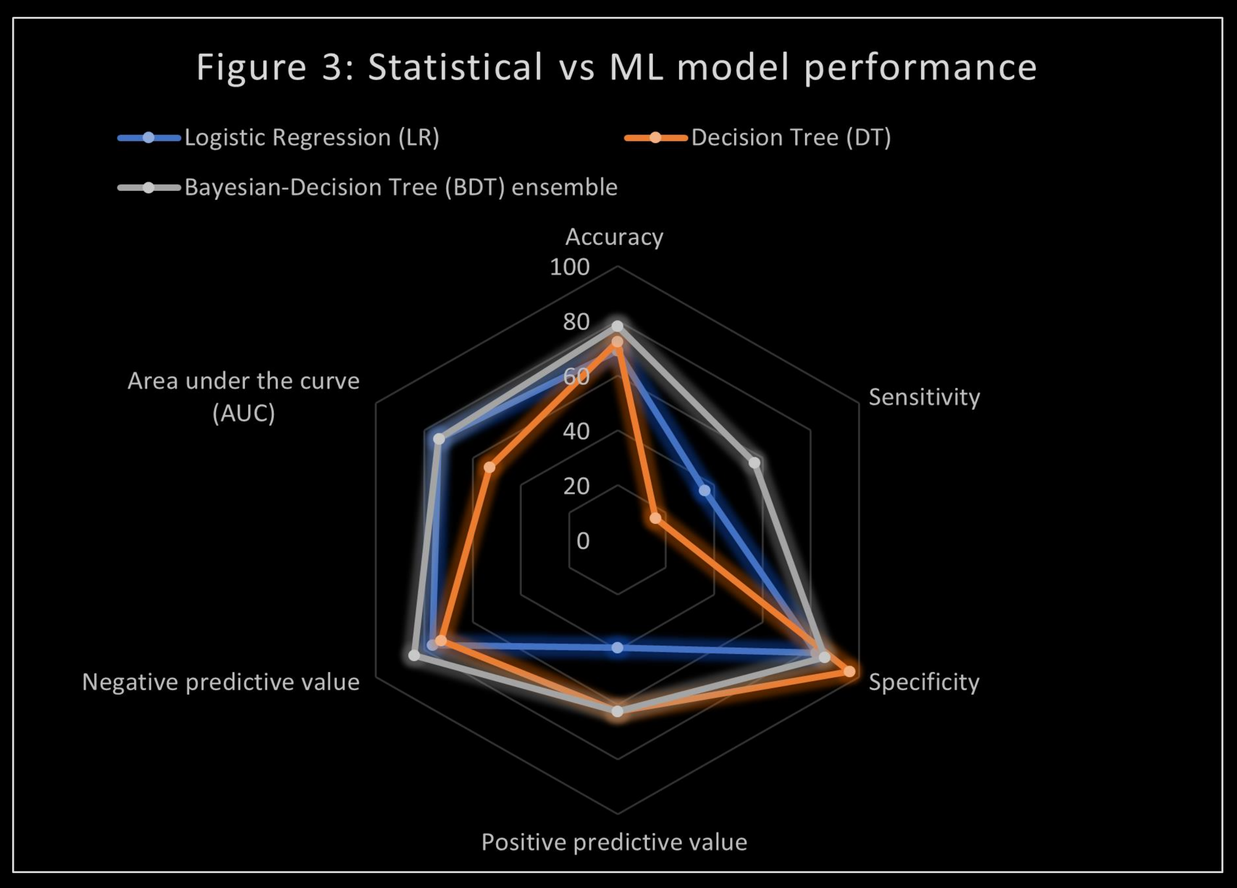
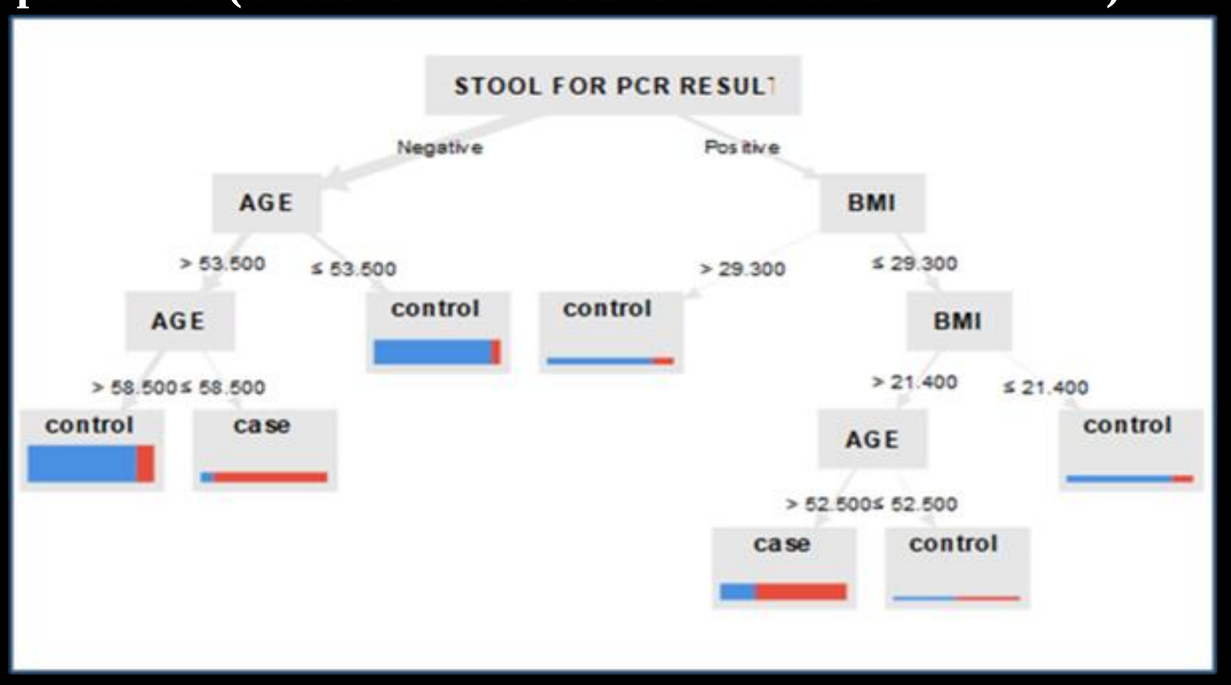


Figure 4 : BDT ensemble early detection algorithm showing stool PCR for SGG, age and BMI as the important CRC predictors (dominant red horizontal bar indicates a case)



1. **Study proved relevance of stool PCR for SGG as a potential screening marker (1).**
2. **Simultaneous testing with iFOBT achieve gains in sensitivity that can improve early CRC detection rates (2).**
3. **BDT algorithm (boosted accuracy to 78%) provided clues that a positive stool PCR patient should be further risk-stratified into normal and overweight BMI and age more than 53 years old to be advised for early colonoscopy.**

CONCLUSION

Ensemble ML model incorporating SGG infection was superior to other models in predicting CRC. Simultaneous SGG screening together with iFOBT for early CRC detection holds potential but needs longitudinal study validation.

ACKNOWLEDGEMENTS

Ministry of Higher Education for the research grant (FRGS/1/2019/SKK05/UIAM/03/1).

CONTACT
Edre Mohammad Aidid, DrPH
Department of Community Medicine, IIUM
Email : edreaidid@iium.edu.my
Office No : +6095704605

MAIN REFERENCES

1. Rezasoltani S et al. The association between fecal microbiota and different types of colorectal polyp as precursors of colorectal cancer.
2. Corredoira-Sánchez J et al. Association between bacteremia due to Streptococcus gallolyticus subsp. gallolyticus (Streptococcus bovis I) and colorectal neoplasia: A case-control study.
3. Arivazhagan B et al. Customer Churn Prediction Model Using Regression with Bayesian Boosting Technique in Data Mining.